

Master Thesis in Applied Information Technology

Comparative Analysis of On-Line Analytical Processing Tools

Sundeep Reddy Maddi and Vaseem Khan
Göteborg, Sweden 2007



IT University
of Göteborg

CHALMERS | GÖTEBORGS UNIVERSITET

Software Engineering and Management



Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

REPORT NO. 2007:39

Comparative Analysis of On-Line Analytical Processing Tools

Sundeep Reddy Maddi and Vaseem Khan



Department of Applied Information Technology
IT UNIVERSITY OF GÖTEBORG
GÖTEBORG UNIVERSITY AND CHALMERS UNIVERSITY OF
TECHNOLOGY
Göteborg, Sweden 2007

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

Acknowledgements:

The accomplishment of this thesis project has been made possible with the direct or indirect involvement of people around IT University of Göteborg. Especially this thesis is a result of brilliant efforts of our supervisor Elisabeth Saalman. We would like to take this opportunity to thank our supervisor for believing on us that we have ability to tackle the situations and taking out time from her busy schedules for supporting us.

We would like to extend our thanks to Thomas Lundquist who continuously motivated us with all kind of support to start this thesis work and we are proud to be the students of Master of Software Engineering and Management program from IT University of Göteborg. We would like to take this opportunity to thank each and every member of the staff for influencing us in choosing our carrier path in right direction.

We are grateful to our colleagues and friends for the nice, purposeful time we had spend with them. We want to say thanks to our parents who never doubted in our efforts.

Sundeep Reddy Maddi and Vaseem khan.

April 2007, Göteborg.

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

Comparative Analysis of On-Line Analytical Processing Tools

Sundeep Reddy Maddi and Vaseem Khan

Department of Applied Information Technology

IT University of Göteborg

Göteborg University and Chalmers University of Technology

SUMMARY

Data Warehouse applications are designed basically to provide the business communities with accurate and consolidated information. The objective of data warehousing applications is not just collecting data and reporting against it, but for analyzing this it requires technical and business expertise tools.

To achieve business intelligence it requires proper tools to be selected. The most commonly used Business Intelligence (BI) technologies are On-Line Analytical Processing (OLAP) and Reporting tools for analyzing the data and to make tactical decision for the better performance of the corporate, and more over to provide quick and fast access to end user request.

This thesis focuses on data warehousing concepts and theories involved with it. The comparative analysis of commonly used OLAP tools is also presented in the thesis.

This report is written in English

Key Words: Data Warehousing, Business Intelligence, OLAP, Cognos, BO (Business Objects).

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

Table of Contents

1. INTRODUCTION	1
1.1 PURPOSE	1
1.2 SCOPE	1
1.3 GLOSSARY	1
2. THEORETICAL FRAMEWORK	3
2.1 DATA WAREHOUSE ARCHITECTURE	3
2.2 ETL PROCESS IN DATA WAREHOUSING	6
2.3 DATA WAREHOUSE SCHEMA MODELS	7
2.4 BUSINESS INTELLIGENCE	8
2.5 OLAP	9
3. METHODS	16
3.1 RESEARCH METHOD	16
3.2 REPORTING TOOLS	17
3.3 DATABASE	29
3.4 PRACTICAL APPROACH	30
3.5 ANALYSES METHOD	30
4. RESULTS	31
4.1 COGNOS	31
4.2 BUSINESS OBJECTS	32
5. DISCUSSION	34
6. CONCLUSIONS	39
7. REFERENCES	40

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

1. Introduction

Data Warehouse is not a single repository system but the process of building decision support systems and knowledge based applications to support decision making and long term business plans. Data Warehouse applications are designed basically to provide the business communities with accurate and consolidated information. The objective of Data Warehousing applications is not just collecting data and reporting against it, but for analyzing this, it requires technical and business expertise tools.

In the following sections we discuss about the purpose and scope of the thesis work. Moreover in this chapter we discuss about Data warehouse environment, On-Line Analytical Processing (OLAP) and Business Intelligence (BI) tools

1.1 Purpose

Business users and managers have difficulties in choosing appropriate BI tool, since a lot of alternatives are available in market. Also many research studies and needs have motivated to the strategic importance of OLAP and multidimensional analysis, which motivated us in exploring reporting functionality of different OLAP tools, and to also review available literature about data warehousing, OLAP, BI and OLAP tools to help business managers in choosing appropriate tool for data reporting and analysis.

It can be narrated in research question as: what is the best available tool for business intelligence.

1.2 Scope

The scope of this report is to describe about the work carried out to achieve the research aim, this report discusses about the Data warehousing environment and its architecture. We are considering different OLAP tools to analyze and compare them.

1.3 Glossary

The glossary (see Table 1) includes abbreviations used in this document.

Abbreviation	Explanation
ETL	Extraction, Transformation, and Loading
OLAP	On-Line Analytical Processing
BI	Business Intelligence
SAS	Statistical Analysis System
CPM	Corporate Performance Management
OLTP	On-Line Transaction Processing
ERP	Enterprise Resource Planning

Table 1: Abbreviations and Explanations

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

2. Theoretical framework

“A data warehouse is a collection of data gathered and organized so that it can easily be analyzed, extracted, synthesized, and otherwise is used for the purpose of further understanding the data. It may be contrasted with data that is gathered to meet immediate business objectives such as order and payment transactions, although this data would also usually become part of a data warehouse”. [1]

The main objective of a data warehouse is "to help people make better business decisions". There is no point in simply accessing data. The point is "really using data". Data warehouses have an intrinsic "breaking-through-barriers nature" and provide comprehensive analytical capabilities that enable anyone to make better data based decisions - instead of basing decisions on bits and pieces of data gathered from several computer applications[2].

Data warehouse is organized and designed to help us in analyzing the data. The data warehouse provides access to difficult environments of an enterprise data. When we build a warehouse of a company's accounts data then the warehouse concentrates on a single subject i.e. accounts of that company.

The data in the warehouse must be integrated and consistent. The word “one version of truth” is often possible only after much discussion and debate about the terms used within the organization. When we are loading the data in a ware house from different sources we should avoid the conflicts such as naming, entities or attributes of an entity during the process of transforming source data.

Many of the figures and numbers that managers receive have little meaning unless compared to historical figures. For example, reports that compare the company's performance now against its performance last year are quite common. The operational systems will not be able to meet this kind of information need, and for a good reason. A data warehouse should be used to record that past accurately, leaving the OLTP (On-Line Transaction Processing) system free to focus on correctly recording current transactions and balances. Actual historical values are not stored on the operational system nor derived by adding or subtracting transactions values against the latest balance.

To meet the business users' needs information can be slice and dice through warehouse data with any means and ability. Dynamic reports allow users to view warehouse data from different angles, at different levels of detail. The ready availability of different data views also improves business analysis by reducing the time and effort to collect format and filter information from data.

2.1 Data Warehouse Architecture

Architecture serves as a blue print of a Data warehouse. By developing data warehouse architecture [21] we can provide:

- Communication tool: This shows the clear picture of the components working together to the members of the data warehouse Group.
- Learning tool: helps to learn more about a new system avoiding trial and error approach.
- Ensuring correct project deadlines.

The figure 1 shows the simple architecture of a Data warehouse.

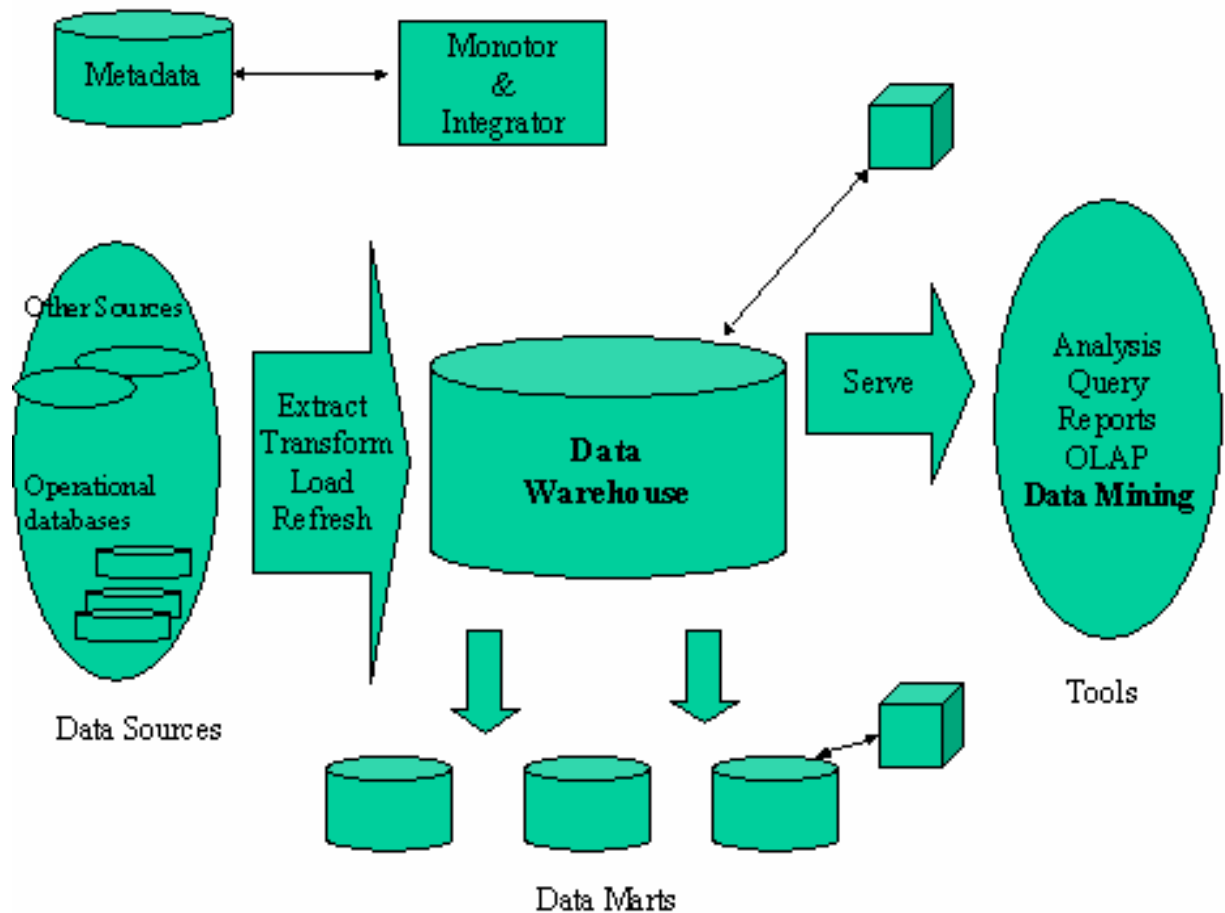


Figure 1: Data warehouse Architecture [3]

In the architecture diagram we have operational data bases at the bottom level with heterogeneous and distributed information. This information is extracted, transform and analyzed by ETL tools to load into Data warehouse. In the Data warehouse the data is stored in relational or multidimensional forms. The Data marts are small chunks of data extracted from the data warehouse for further study and to solve business problems. In the final form the data is analyzed according to the end users request using reporting tools

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

There are several advantages of data warehousing and here we are discussing some of the advantages [19] :

"The warehouse has enabled employee to shift their time from collecting information to analyzing it and that helps the company make better business decisions" a data warehouse turns raw information into a useful analytical tool for business decision-making. Most companies want to get the information or transaction processing quickly in order to make a decision-making. If companies are still using traditional online transaction processing systems, it will take longer time to get the information that needed. As a result, the decision-making will be made longer, and the companies will lose time and money. Data warehouse also makes the transaction processing easier.

In order to minimize inconsistent reports and provide the capability for data sharing, the companies should provide a database technology that is required to write and maintain queries and reports. A data warehouse is a central repository which provides all the metrics necessary to support decision-making throughout the queries and reports. Queries and reports make the management processing efficient.

A typical data warehouse objective is to store all the information including past data and the current data. Historical data is seldom kept on the operational systems; even if it is kept, it is rarely found in three or five years of history in one file.

The primary work of all data warehouses is to integrate data, because it is a primary deficiency in current decision support. Another reason to integrate data is that, data in one file is updated at different time period than that in another file this effects the level of granularity.

Even though corporate have advantages of using data warehouse, but there are some disadvantages also. Data warehouse system is too expensive. According to Phil Blackwood in 2000 "The average cost of data warehouse systems valued at \$1.8 million" [23]. This limits small companies to buy data warehouse system. As a result, only big companies can afford to buy it. It means that not all companies have proper system to store data and transaction system databases. Furthermore, because small companies do not have data warehouse, then it causes difficulty for small companies to store data and information in the system that may cause small companies to organize the data as one of the requirement for the company to grow.

Moreover, data warehouse is a very complex system. Data warehousing can complicate business process significantly. For example, small change in the transaction processing system may have major impacts on all transaction processing system. Sometimes, adding, deleting, or changing the data and transaction can cause time consuming. The administrator need to control and check the correctness of changing transaction in order to impact on other transaction. Therefore, complexity of data warehouse prevents the companies from changing the data or transaction that are necessary to make.

2.2 ETL process in data warehousing

The process of extracting data from source systems and bringing it into the data warehouse is commonly called ETL which stands for extraction, transformation, and loading [4]. To facilitate the business analysis, we need to extract data from one or more operational systems and copy into data warehouse regularly. The term ETL means extracting data from outside sources and transforming it to fit into business needs and in the end loading into the data warehouse.

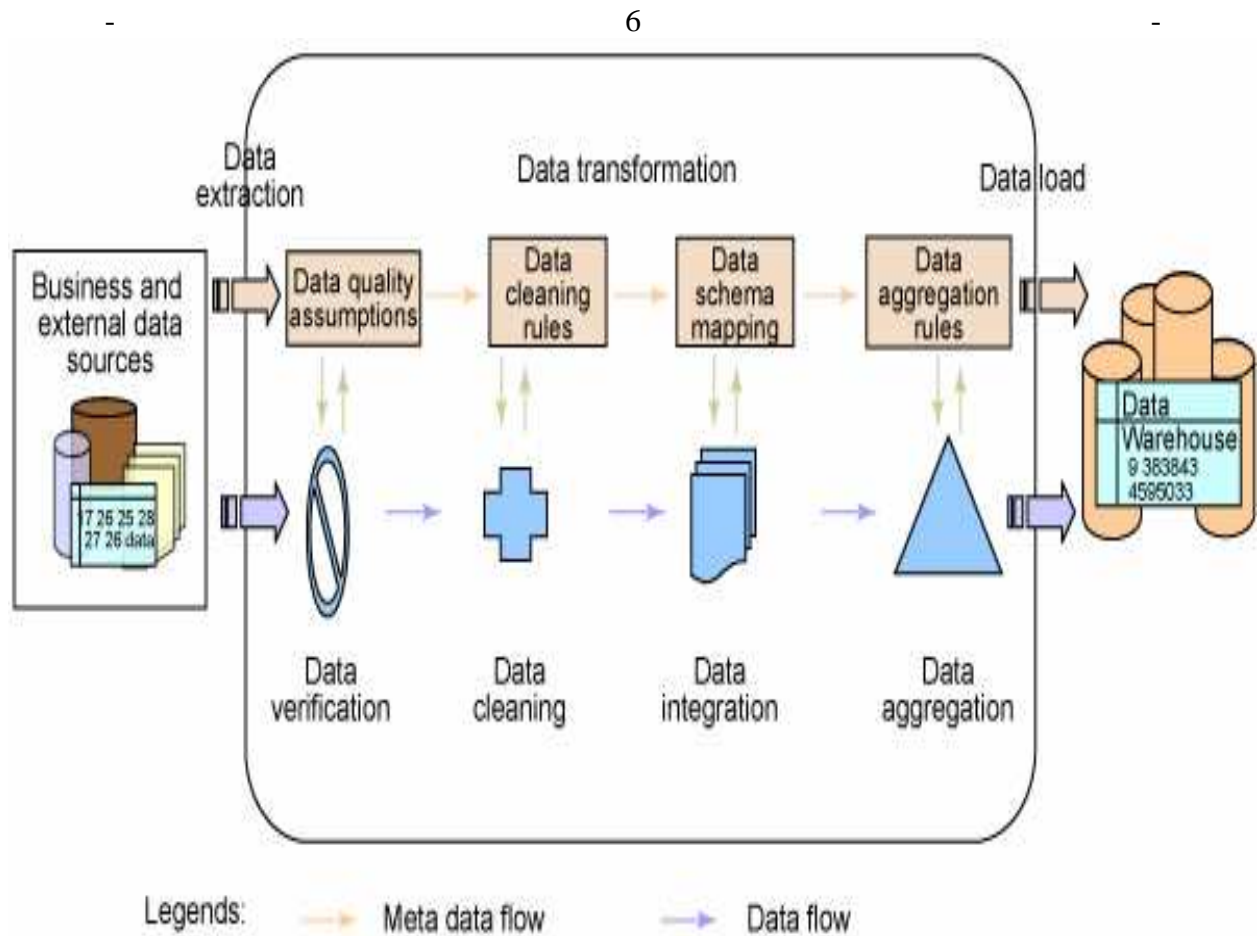


Figure 2: ETL data transformation process [17]

In the ETL process, the first step is to extract data from different source systems. Each source system may use different format and different type of data organization. The commonly used data sources format are relational databases [20] and flat files. In this step data is extracted in a format for transformation processing. In the next step of

transform phase a series of rules are applied to extracted data to derive the data to be loaded. Manipulation of data is done in different transformation types depending on the data sources. In the last phase data is loaded into data warehouse depending on the requirements of the company. This process may vary from organization to organization in some warehouse may overwrite the old information with the new data and in some systems may maintain a history and audit trail of all changes of the data.

2.3 Data Warehouse Schema Models

A collection of database objects, including tables, views, indexes, and a synonym is called a schema. There are different ways of arranging schema objects in the schema models designed for data warehousing. Data warehouse schema model used in designing our database is a star schema. Following are the commonly used schema models for data warehouses.

2.3.1 Star Schema Model

A schema is a collection of data base objects, tables, views and indexes. There are different types of schema models designed for data warehousing depending on the variety of ways of arranging schema objects. The most common data warehouse schema is star schema; this is perhaps the simplest data warehouse schema. This is called star schema because the entity relation ship diagram of this schema resembles a star in the center which consists of large fact table [18] and end points are the dimension tables. A single star schema is characterized by one or more fact tables which contain the quantitative data of the data warehouse and the dimensional tables are number of smaller tables which contains information about the entries for a particular attribute in the fact table this data is usually a descriptive data.

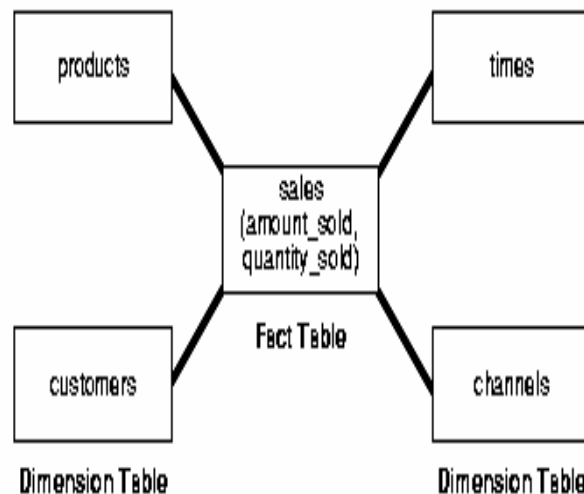


Figure 3: Simple Star Schema [7]

To join a fact table and dimensional tables a star query is used. In a simple star schema the primary key of a fact table consists of one or more foreign keys joined to different dimensional tables but the dimensional tables are not joined to each other.

Star schemas can provide a direct mapping between entities being analyzed by end users and schema design, this schema is widely supported by a large number of business intelligence tools; it gives good performance for any typical queries.

2.3.1.1 Snowflake schemas

This schema is more complex data warehouse model than the star schema. This is also a type of star schema but it is called snowflake schema because the diagram of the schema resembles a snowflake. To eliminate the redundancy dimension data has been grouped into multiple tables instead of one large table in snowflake schema to achieve normalization.

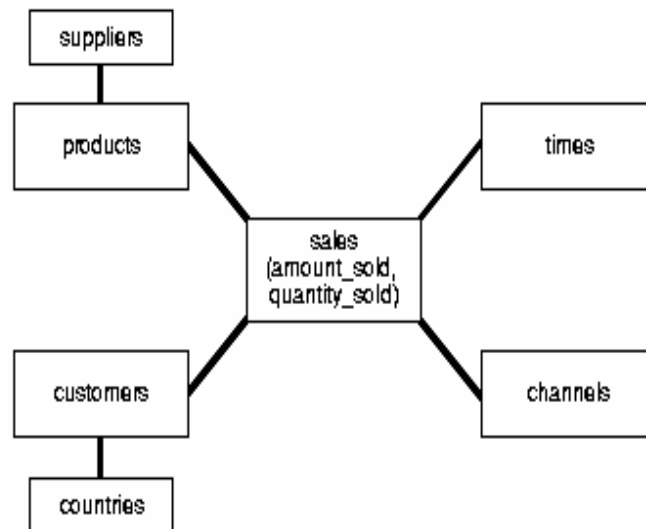


Figure 4: Snowflake Schema [7]

2.4 Business Intelligence

Business intelligence concept is popular and powerful technologies used to turn data into meaningful information. Using business intelligence applications, bulk amount of data originating in many different formats (ex: spread sheets, relational database and web logs) can be brought single and presented to key business analyst quickly and more accurately. Thus the intelligent information that is easily understood, the business analyst can effect the change and develop plans to achieve profits. For the organization the pace of the business decisions are increasing day by day to stand tuff in the growing market competition. Due to increase in data at our disposal we often find it difficult to understand what the data means. Technologies advancement have never been ending and we need to the technology alone is not the solution for the problem. To move forward the

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

companies' strategies they need people who can make tactical decisions and solve business problem.

To make faster and better decisions, business managers and executives require information in the form of useful data that is available and easy to analyze. The gap that exists between the business information and the people need and the raw data collected by companies are increasing tremendously. In order to over come this gap companies has to recognize how the bulk volumes of data could be used for business analysis and should understand how the available raw data be converted into useful information. Through a series of processes that begin with OLTP systems and end with BI systems can bridge the gap between the raw data and business information. OLAP systems allows adhoc analysis and dice data available in different dimensions, pivoting rows and columns and drilling down through out hierarchies. The standard OLAP capabilities allow the end user to access quick and fast answer questions.

Business intelligence term is used to characterize a range of technologies, software platforms, specific application and processes. Since BI is all round concepts, it can be featured into three different perspectives

- Using a reasonable approach to management
- Data can be converted into information
- Making better and faster decisions

Most of the vendors have front end presentation layer which allows users to create reports. Users should check these functionalities while evaluating reporting tools:

- Data source connecting capabilities
- Scheduling and distribution capabilities
- Security features
- Customization
- Export capabilities

2.5 OLAP

On-line Analytical Processing (OLAP) is a category of software technology that enables analysts, managers and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user. [5].

OLAP functionality is characterized by dynamic and multidimensional analysis of an enterprise data for end user activities. These activities include a wide variety of views or multidimensional conceptual views of the data by supporting hierarchies. It is easy to understand the data mart designed for OLAP analysis by end user. An OLAP server helps

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

the user to generate fast reports regardless of database size and complexity and users are allowed to define new adhoc calculations in any desired way.

OLAP is implemented in multi-user and client server modes regardless of database size and complexity and offers quick response to queries. This helps users to go through historical and projected data in different data model scenarios and these are achieved only through use of OLAP server.

In the traditional script- based analytical reports are usually generated by extracting data from two-dimensional relational database tables. These two dimensional reports are formatted and delivered to users but these reports have disadvantages such as:

- These reports cannot be easily modified; they are predefined, built-in reporting applications
- Depending on the business reporting requirements only IT professionals can modify the report or build new report.
- Business data is in different place from the reporting user.

On the other hand, OLAP reporting integrates the complex issues of an enterprise data structures, procedures and algorithms into its dimensional data structures. These are presented to end users in easy to understand dimensional information views in a very instinctive way. Users can easily navigate the predefined reports and read the business data to build new adhoc reports with small support from IT professionals. By this way users become aware of existence business dimensions and get near to business data to answer the queries. Following are some benefits of OLAP [22]:

- OLAP tools can improve the productivity of the whole organization by focusing on what is essential for its growth, and by transferring the responsibility of the analysis to the operational parts of the organization.
- Developers can benefit from the OLAP software. By using software specifically designed for OLAP, developers can deliver applications to the business users faster, providing better service that in turn allows the developers to build more applications.
- Advantage of using OLAP systems is that such systems are separate from the On-line transaction processing (OLTP) systems that feed the data warehouse, OLTP systems performance will improve due to the reduced network traffic and elimination of long queries to the OLTP database.
- OLAP enables the organization as a whole to respond more quickly to market demands.

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

OLAP tools and applications can be used by a variety of organizational division's sales; marketing, finance and manufacturing are some of them.

The key features of OLAP applications are Multidimensional Views, Complex calculations and time intelligence.

Multidimensional Views: Multidimensional views represent the actual business model and they are limited to less than three dimensions. Managers usually look at the financial data by scenario that is actual versus budget and at sales by product and time. This view of data provides more than the ability to slice and dice, gives flexible access to information for analytical processing. Database design should not influence which operations to be performed on a dimension and managers must be able to analyze data at any level with functionality and ease.

Complex Calculations: OLAP database can be verified by its ability to perform complex calculations; these databases must be able to do more than simple aggregation. For example calculation of percentage of total from a top down perspective. While doing aggregation it should follow hierarchy analyses rather than simple data roll-ups. The key performance of a database involves algebraic equations, making algorithms such as moving averages and percentage growth of a given company.

Time intelligence: Time is an essential element of almost any analytical applications. A true OLAP system must understand the sequential nature of time and always judge over time for business performance for example this day with that of the last day of the last year.

OLAP applications can support many types of business area and to achieve more profits on investment. Here are the few examples of analytical applications.[6]

- *Accounting:* Forecasting, budgeting, cost and profitability analyses, and consolidation
- *Human Resources:* Skills consolidation, labour scheduling and optimization
- *Distribution:* Scheduling and optimization
- *Sales Force:* Automation. Cross-selling and territory analyses
- *Marketing:* Churn and market-based analyses
- *Retailing:* Site location and demographic analyses
- *Manufacturing:* Demand planning and forecasting
- *Health Care:* Outcomes analysis
- *Financial Services:* Risk assessment and management

OLAP applications have been most commonly used in the financial and marketing areas and their uses do extend to other functions. Data rich industries have been the most

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

typical users (consumer goods, retail, financial services and transport) for the obvious reason that they had large quantities of good quality internal and external data available, to which they needed to add value. However, there is also scope to use OLAP technology in other industries. The applications will often be smaller, because of the lower volumes of data available.

2.5.1 OLAP System Components

An OLAP system consists of multiple components. A top level view of the OLAP system includes a Data Source, OLAP server and a Client. In the data source we have source of data to be analyzed and this data is transferred into OLAP server, where it is organized and arranged to provide short queries. The last component client is the user interface to the OLAP server. [9]

Data Source: A Data source is an OLAP system that supplies the data to be analyzed. The data source could be a data warehouse, a legacy database, a collection of spread sheet or combination, these depends on the use of OLAP product. It requires that all source data is stored in a specific format in a database is problematic for database Administrators. Administrators should check that the OLAP product works properly with combination of number of data sources. The products which allow data extraction from wide variety of data sources are useful and flexible for the administrators and users.

OLAP Server: The back end of the OLAP system is the OLAP server which does all the work and where data that is retrieved is stored. The architecture of the server depends on different philosophies; the major future of OLAP product is that, server uses a multi-dimensional database (MDDB) or a relational database (RDB) to store the data.

Client: The client is what is spent to view and change the data in the database. A client can be a spreadsheet with OLAP features or it can be a simple report viewer or it can be more complicated application for data manipulation. For example World Wide Web is the newest of the client. The server is the mainstay of an OLAP solution which provides a solid foundation for quick data manipulation. The front end is also important for the user to get benefit and most of the vendors focus on building front end applications.

Query tool / Report writer: A query tool or report writer provide a modest tool to access OLAP data these tool allow users to create report by dragging and dropping objects with a graphical interface. A traditional report writer allows users to easily produce formatted reports.

2.5.2 OLAP Architectures

OLAP products are classified as either multidimensional OLAP or relational OLAP based on the underlying architecture of the system. When combining the strength of ROLAP and the advantages of MOLAP, a third option for implementing OLAP can be created hybrid OLAP (HOLAP). In some cases it may also be reasonable not to build a separate database for OLAP but to work directly on the OLTP system in a OLAP fashion which is

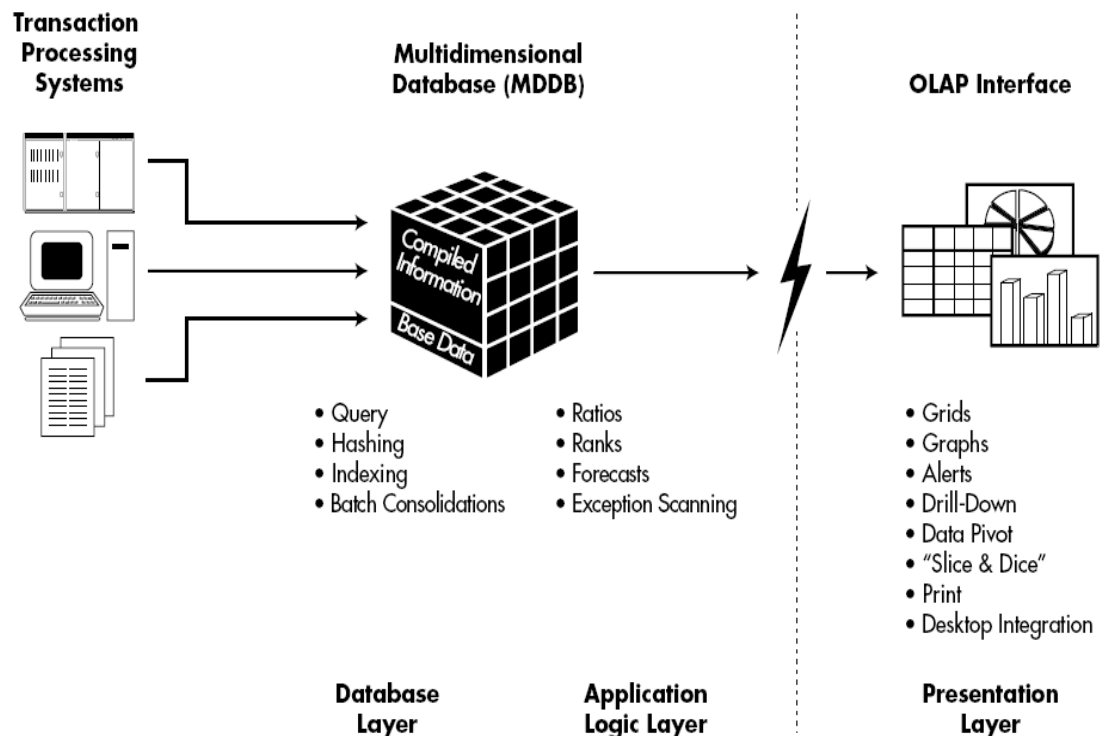
Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

called Virtual OLAP. This section describes some of the architecture which are commonly used.

2.5.2.1 Multidimensional OLAP Architecture

Multidimensional OLAP (MD-OLAP) uses a proprietary MDDB to provide OLAP analyses. The purpose of using MDDB is that it can efficiently store data which are by nature multidimensional to provide a means of quick querying of the database. Below figure5 [11] gives the overview of the general MD-OLAP architecture. Information from all different data sources is transferred into multidimensional database through a series of batch processing. Once the operation of extract, transfer of data into the MDDB is over, now the process of performing serious of calculations in batch routines is done to fill the MDDB array structure. For example all the state revenue figures for all the stores would be added to fill state cells in the database. When the array structure in the database is filled, hashing algorithms and indices are created to improve the uery accessing time.

Once the process is finished, the MDDB is ready for use. Users can know request for OLAP reports through interface, and the stored data is retrieved through the application logic layer. MD-OLAP architecture is a complicated model. It reads the precompiled data, and has limited capacity to dynamically aggregate or difficulties in business metrics that have not been calculated and stored.



Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

Figure 5: Multidimensional OLAP Architecture [11]

MD-OLAP is a two-tier, client- server architecture. In this design, MDDB serves as both the database and application layers were MDDB system is responsible for data storage, access and retrieval processes in the database layer. In the application layer MDDB executes all OLAP requests. End user requests and OLAP analyses is carried out in the presentation layer. Multiple databases are accessed by multiple users in this client/server architecture of MOLAP.

2.5.2.2 Relational OLAP architecture

ROLAP main function is to access data stored in data warehouse to have OLAP analyses. The principle idea of ROLAP is that OLAP abilities are best provided against the relational database. When the data model for the data warehouse is defined, data is loaded from the transaction processing into database. Database practices are run to collect the data as required by the data model. Indices are created to optimize query access time period. Clients submit multidimensional analyses to the ROLAP engine, which dynamically transfer the end user request into SQL execution plans. Now SQL is submitted for the relational database processing where relational query results are cross-tabulated and a multidimensional result is generated and is send back to the end user. ROLAP is three-tier, client / server architecture where database layer utilizes relational databases for data storage, access and retrieval processes. ROLAP is a application layer which executes multidimensional reports from multiple end users. OLAP analyses are done by the users when ROLAP engine join together with variety of presentation layers.

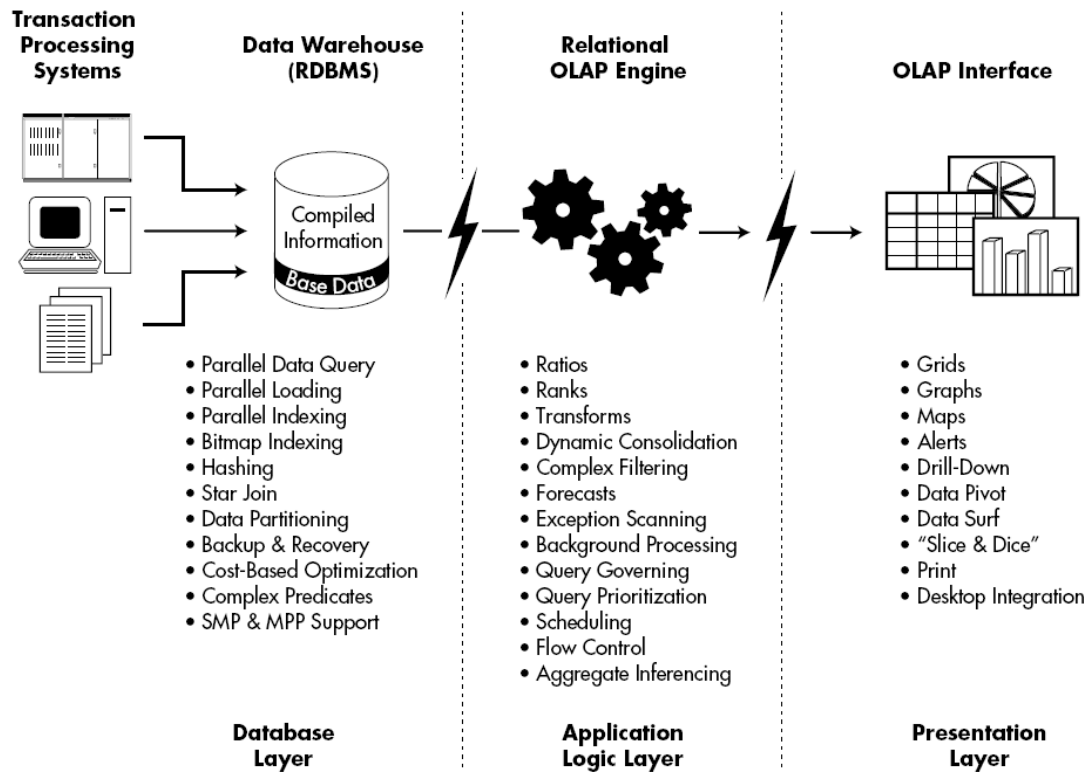


Figure 6: Relational OLAP Architecture [11]

3. Methods

The Methodology chapter explains and justify the way of conducting the research study. The basic idea of the methodology is to easily understand how the work is proceeded. The most important mission for methodology is to clarify how we use the method, solution of problem, study plans, techniques and the study areas make up the parts of an integrated whole.

3.1 RESEARCH METHOD

The project is divided in to two parts; the first part is to conduct literature study about data warehousing, business intelligence, and OLAP and ETL process. Because it is necessary to understand background knowledge associated with business problem, before going deep examining respective practical methods or tools. The second phase is to explore reporting functions of business intelligence tools like Cognos, Business Objects, SAS and Micro strategy.

For the first purpose a thorough literature study has been conducted. Literature sources like IEEE explorer, ACM library and Chalmers digital library have been used.

The second purpose was to explore reporting function of BI tools. To accomplish that, explorative research approach was followed. The tools were explored and reports were generated and at the end tools were analyzed against selected factors.

The approach used in the overall effort is shown in the following figure.

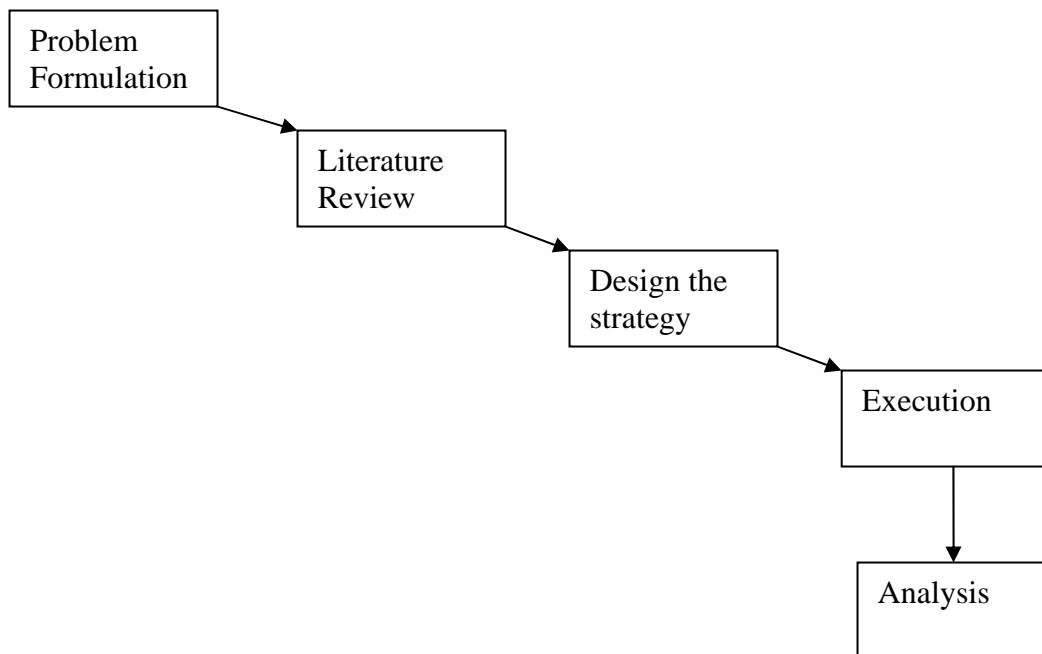


Figure 7: Research methodology

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

The strategy design phase will serve the purpose of creating factors to be analyzed at the end, planning how and which tools will be explored. In the execution phase tools were installed, configured for walk through. At the end analysis were performed by correlated results of execution phase with selected factors.

3.2 Reporting TOOLS

The most common applications of reporting tools are business operations reporting and dashboards. Depending on the organization data is analyzed and presented.

In the thesis we have implemented different available OLAP tools. These are commonly used by advanced users to look at the data from multiple dimensions. The main purpose of OLAP tools are slicing and dicing of data, for these they require strong metadata layer and front end flexibility. In the OLAP tools, we have choose the most commonly used tools such as Cognos, SAS, Micro strategy and B.O

We are analyzing some of the OLAP tools that are mostly used in the market. The main purpose is to study its environment and more over we want to implement these tools in our thesis work.

3.2.1 Cognos

Cognos is the one of the leading provider of enterprise planning, business intelligence and corporate performance management software. It uses the software for decision support, data mining, graphs and reporting from different data sources.

"CPM is an umbrella term that describes all of the processes, methodologies, metrics and systems needed to measure and manage the performance of an organization." [12]

Organizations performance is increased by understanding the values of CPM. Many organizations have understood how CPM can solve the fundamental business questions such by delivering scorecards, business intelligence, and planning and consolidation technologies.

Reporting is one of the products, which are capable to provide business intelligence solutions to the organization. Multilingual reporting can be derived from a single metadata layer in reporting. Dash boards are generated by BI Reporting; these are used to translate useful information from huge volumes of corporate data into graphs, maps, charts and other graphical elements.

Analysis is a business intelligence product which enables the analysis of information from all dimensions of a business data. It lets us analyze changes over a periodical time. This tool is easy to use and data can be easily deployed.

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

From the help documentation of cognos [8], once we start cognos impromptu we have to establish an ODBC connection with the database

To start impromptu administrator, click on Start -> Databases -> Cognos EP Series 7 -> Cognos Impromptu Administrator.

To create an ODBC connection, from the menu click on Catalog -> Databases (from the figure8 the first window will be opened.). From there click on ODBC -> Select New Database.

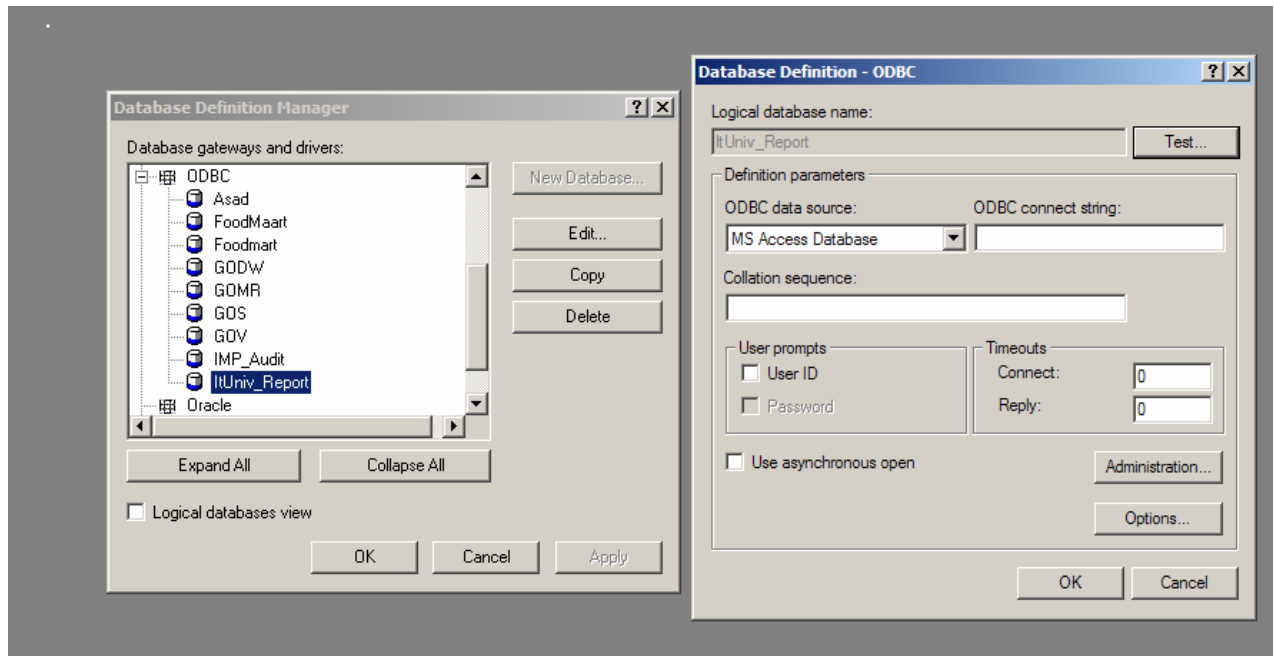


Figure 8: Establishing ODBC connection with Cognos

Select the administrator option. This will open the ODBC administrator window then click on the system DSN. With system DSN, all the users who can access the system can also the same ODBC instead of creating a new one.

Now click insert button and select Microsoft access driver. Then ODBC will ask for the name, description of the ODBC and the location of the database in the system. Provide the details.

This will create a new ODBC connection for the data, the same being shown in the below figure9.

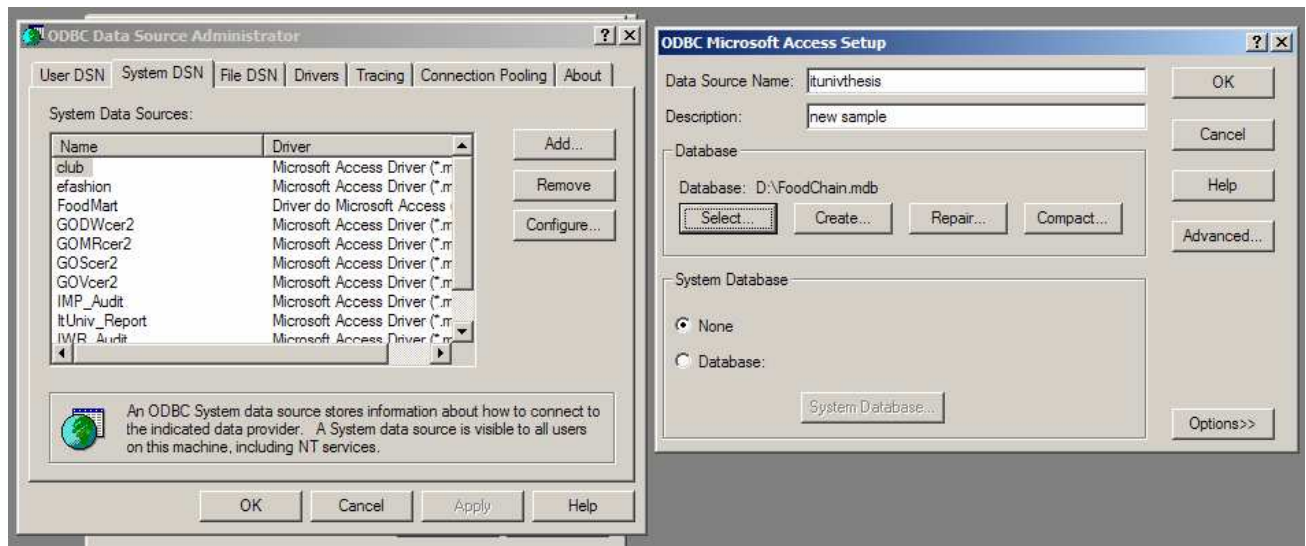


Figure 9: Creation of ODBC Connection

To work with this ODBC in Cognos, firstly, establish a logical connection to the ODBC which in turn will connect to the Database. From the menu select Catalog -> Databases -> ODBC -> New Database Name. This will create a logical database connection to the database. From the figure9, the second window will be opened. Give the logical database name and other details.

For Thesis work, taken, “ITuniv thesis” as the logical database name. Ones we enter the details, the logical database name will be entered into the Cognos database manager window. This will establish a connection to database.

A Catalog is a place, where, tables and other information which is required for the generation of the cube from the database can be stored. The same is explained in the following figure10. Select Catalog -> New, figure10 shown below will be opened.

Enter a database name, description details. The catalog type specifies, whether all can access the catalog, if it is personnel, or shared etc., In Cognos there are four types of catalog types. In the drill down window, select the database name, means enter the logical database which is created in the previous step, which in turn will connect to the database.

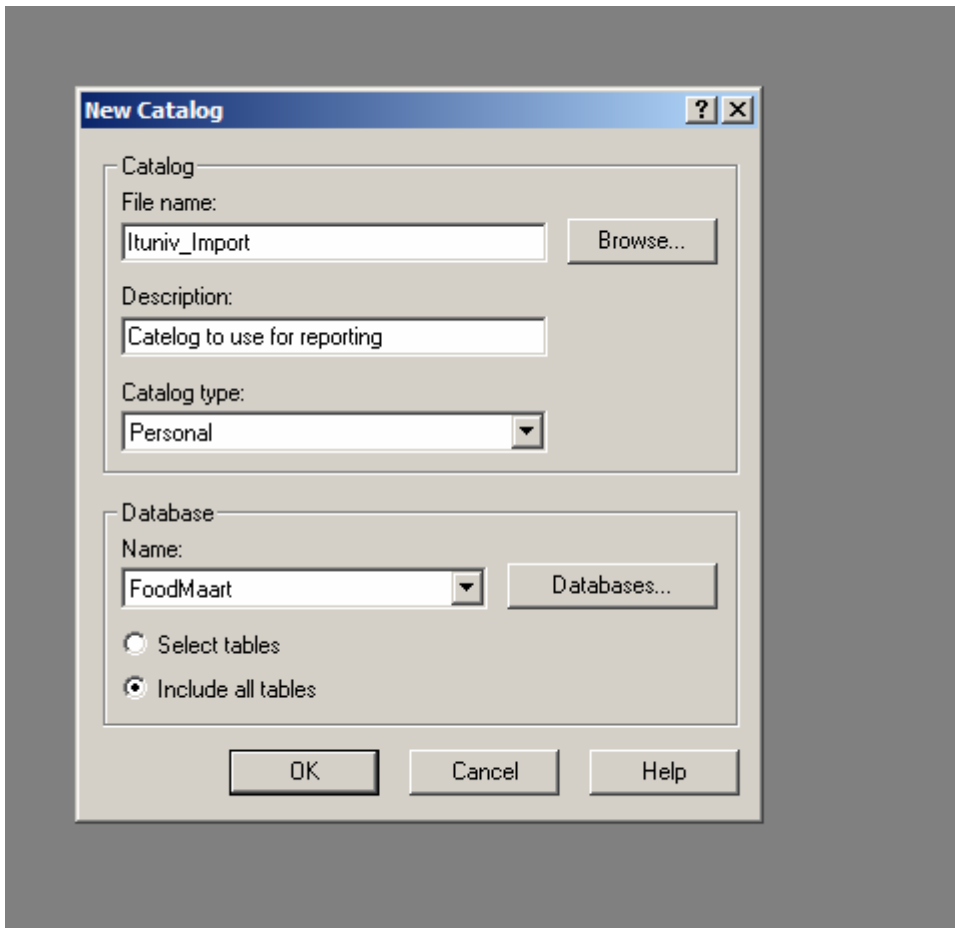


Figure 10: Creating a New Catalog

At last in the option buttons, whether all tables are selected or only part being selected is to be entered, select the option which is convenient. For our Thesis, we have taken all tables. Press ok to complete the general properties of catalog creation process.

Ones we select ok, the next window will be opened to select the tables. All the dimension and fact tables in the created Star-schema will be visible. Double click on each table to know the fields in each table. Select the table if the whole table is required for the purpose or only the fields which are required from each table and click on insert to insert them into the catalog. Then give the database name in the drill down window. Also specify the relationship among the table, for our purpose it is already being created in the process of creation of Schema itself so we have chosen the same to work with. This completes the work of creating a catalog.

Now generate the output in the report format using Cognos impromptu. From the menu select, File -> New option. This is used to create a new report in the Cognos administrator. Ones clicking on new from the file menu, Cognos will show different

formats of reports, which can be generated to display data in different formats. Among the available reports we have selected one report format.

Select the required fields from each table. For generating a report select some fields among the available tables.

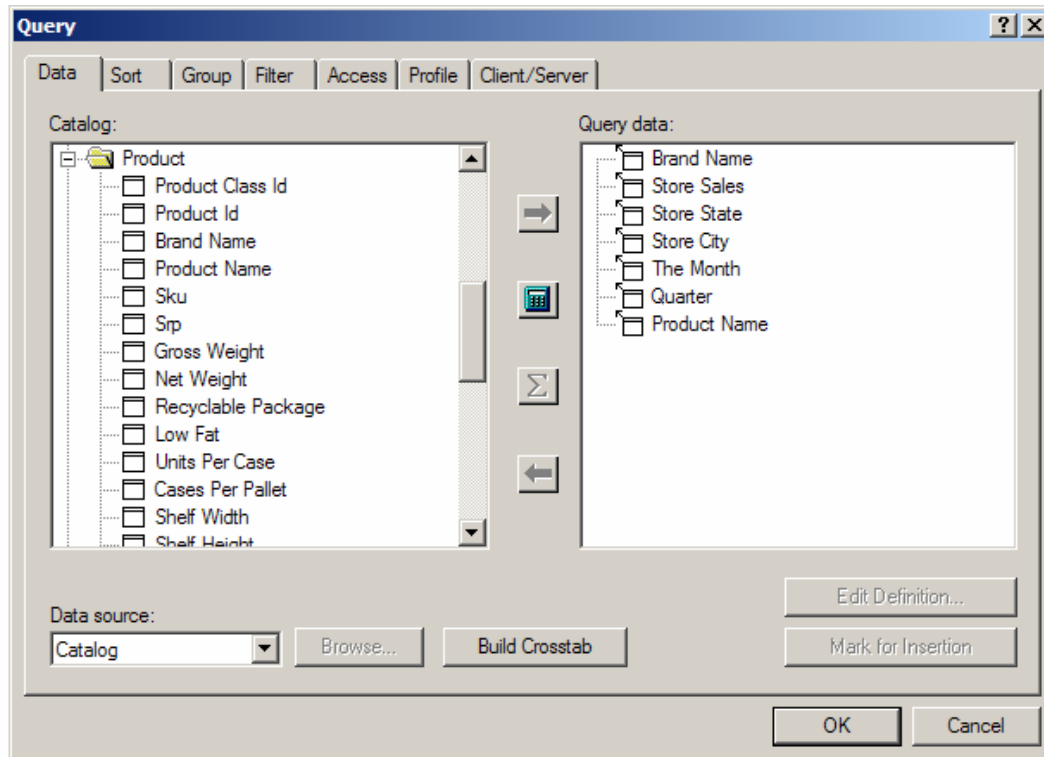


Figure 11: Selecting Fields to make a Query

Each tabs like group, filter etc, which being displayed in figure11 are used to simplify the query and to present to data according to our requirements.

3.2.2 Statistical Analysis System (SAS)

SAS is one of the best providers of business intelligence software which gives us information in any format when we need it. It integrates data from variety of platforms and applications across an enterprise and delivers self service reports and analysis. IT can make better solutions and more informed decisions by spending less time responding to user's request.

SAS provides business solutions for departmental applications through an integrated suite tools. This tools transfers bulk data within organization into useful information. Due to fast and accurate analysis of data help business organization to analyze millions of

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

records and allow new data to flow in an organized form. This allows business organizations to know the changes and to make organized of bulk volumes of available data and more over allows accurate and timely decisions.

SAS business intelligence offers quick business decision making and more ease to generate reports and analysis. Reports are framed in the form were the organizations can easily access it and pull the data from different sources according to the individual needs. SAS reporting tools help in faster and quick availability of information, lower cost on ownership and to deploy critical business intelligence to more people.

SAS provides web based wizards and interfaces and Microsoft office interfaces which make simple for organization for querying and analyze data. Microsoft office environments are used to perform complex analyses thus IT can obtain more strategic objectives.

3.2.3 Micro Strategy

Micro Strategy is also one of the business intelligence tools which give business solutions to users' queries. It allows reporting and analysis of data stored in a relational database.

Reporting is done in Micro Strategy by use of relational database technology. A common unified metadata is provided to support reporting in business intelligence.

Micro strategy supports five styles of BI by:

1. Enterprise reporting is carried out with pixel reports formats for operational reporting and dashboards.
2. Cube analysis is done for limited datasets.
3. Adhoc query and analysis is done by slice and dice OLAP analysis of the database.
4. Statistical analysis and data mining of data is done for correlation analysis, financial analysis.
5. Alerts and Reports are delivered on schedules or event triggers of database.

3.2.4 Business Objects (BO)

Designer is a Business Objects product intended specifically for customers as the universe designer. With Designer, we can create universes, the semantic layer that isolates end users from the technical issues of the database structure. Universes can be dedicated to specific groups of end users in the company or organization.

The Universe Designer in a Company may hold any one of the following positions database administrator, applications manager, applications developer, project manager, etc. In fact, there may be more than one universe designer in our company and the number of universe designers necessary depends, of course, on a company's particular

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

needs or requirements. For example, one universe designer can be appointed per application, project, department or functional area or one universe designer may be sufficient for the entire company. To determine the types of universes appropriate for our organization, the universe designer may interact with other persons such as the Business Objects supervisor, the database administrator, or end users.

The Tasks of the Universe Designer: Universe designer is responsible for Designing, creating, and maintaining universes and distributing the universes.

The Metadata Layer: the Universe designer is responsible for creating universes, the Business Objects products metadata layer. This metadata layer is based on the paradigm of universes, classes, and objects.

Universes: A universe is a business-oriented mapping of the data structure found in databases tables, columns, joins, etc. A universe can represent particular application, system, or group of users. For example, a universe can relate to a department in a company such as financing or accounting. A universe may also pertain to a section within a company or just about any set of organized procedures such as a payroll or inventory system. A universe consists of a set of objects, grouped by classes.

Classes: A class is a logical grouping of objects within a universe. A class reflects a business concept that conveys the category or type of objects. A class can be further divided into subclasses. By using designer, we can define hierarchies of classes and subclasses into a model that best reflects the business concepts of a company. Attendance Information, No of Sick Days, Employee Presence, Department Information these are some typical examples of Class.

Objects: An object is the most sophisticated component in a universe. An object maps to data or a derivation of data in the database. For example, for a product manager, objects might be Product, Life Cycle or Last Date while for a financial analyst, objects might be Profit Margin, Return on Investment, etc.

In multidimensional analysis, objects are qualified as one of three types dimension, detail, or measure. The type dimensions typically relate to a hierarchy of product or time and dimension objects are the parameters for the analysis. In detail objects we can provide a description of a dimension. A measure object gives numeric information by which a dimension object can be measured..

Creating a Universe: To create a basic universe, we have to design a database, so we took the database with Microsoft Access version 2000. This database is used for the thesis. During the first phase, we create the underlying database structure of our universe. This structure includes the tables and columns of a database and the joins by which they are linked. We may need to resolve loops, which may occur in the structure by using aliases or contexts. We can conclude this phase by testing the integrity of the overall structure.

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

From the help documentation of BO [10], we have the steps to create a universe based on a simple relational schema, DESIGNER provides Quick Design, a wizard for creating a basic yet complete universe. We can use the resulting universe immediately, or we can modify the objects and create complex new ones. In this way, we can gradually refine the quality and structure of our universe.

When we start the Designer it displays an overview of the four steps necessary to create a basic universe. It also provides a check box Click here to choose strategies. If we click this check box, we will be able to select the strategies for creating the universe; otherwise, Designer applies the default built-in strategies. In each dialog box that follows, Quick Design prompts we for the information needed to carry out the action.

To move from one dialog box to the next, click the Next button. We can return to the previous dialog box by clicking the Back button. We may end the process and quit Quick Design at any time by clicking the Cancel button. Click the Begin button to start the creation process.

Defining the universe parameters: In this step, we define the universe parameters the universe name and a database connection as indicated in the figure12. We can either create the connection, or select an existing one. To create a connection, click the new button, we can see “add connection screen” this screen leads to establish ODBC drivers to our database to business objects. In this screen we have to select ODBC administration the ODBC Data source screen will appears, In this screen we will create ODBC driver to our data source by giving specific name and the location of our Database. Now the connection was established between the database and Business Objects. By pressing next button will lead us to the Classes and Objects wizard.

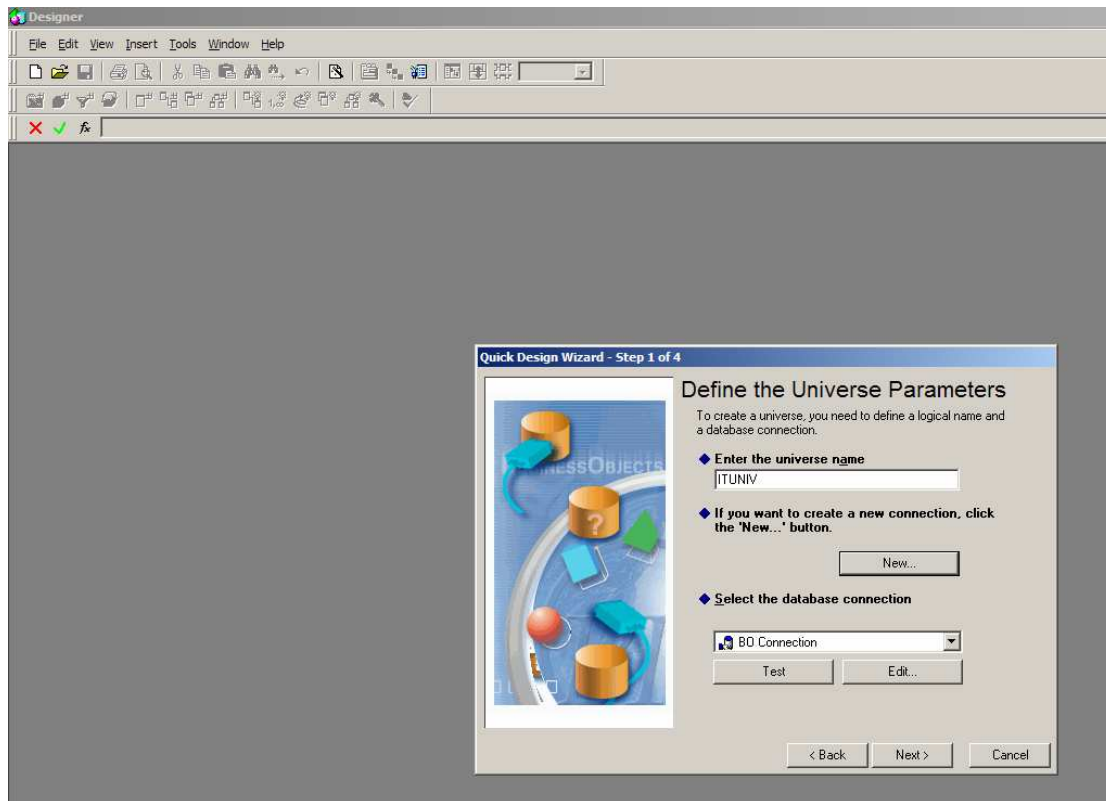


Figure 12: Establishing ODBC Connection to our Data Source in Step 1

Choosing the strategies: If we click the check box for strategies in the welcome screen, Quick Design prompts us to specify strategies for the creation of objects, joins, and tables. A strategy is a script that reads structural information from a database or flat file. This information can pertain to tables, columns, or joins.

Creating the initial classes and objects: Based on the parameters of our database connection, the wizard presents us with a list of database tables and columns. We create the initial classes and objects by selecting tables and columns from the left pane, and adding them to the Universe classes and objects pane on the right. In figure13 by default, the left pane shows only the names of the tables. To view the columns of any table, double click on the table name. To view the data values of any table or column, click it and then click the View Values button.

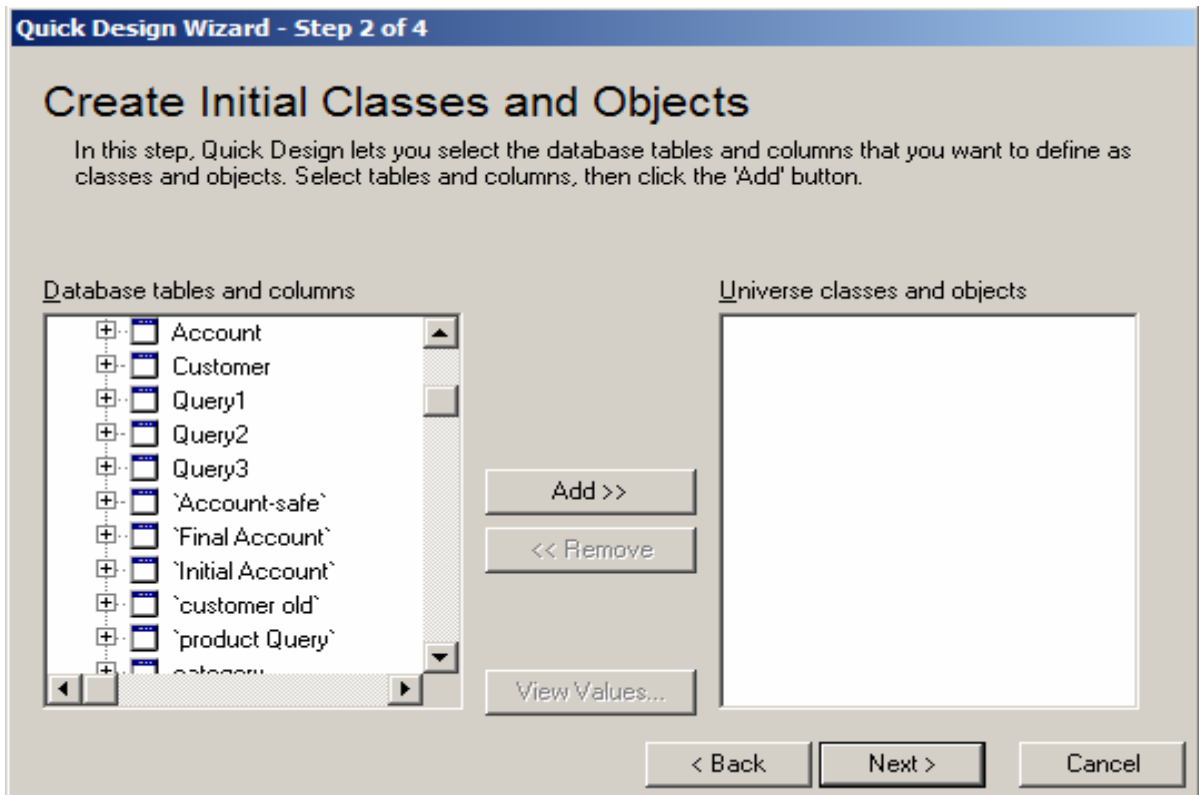


Figure 13: Creating Classes and Objects Wizard

To select one table, click the table, and then click the “Add” button. To select several contiguous tables, hold down the Shift key, then click the first table and last table. When we insert a table, DESIGNER includes all of its columns. In the right pane, the names of classes are displayed beside a folder icon. Click the plus sign (+) beside the class name to view the objects. We can rename a class or object by double-clicking it and entering a new name in the dialog box.

Creating measure objects: A measure object is derived from an aggregate function Count, Sum, Minimum, or Maximum. This type of object provides numeric information. To create a measure object, click the appropriate object in the left pane, and then click the aggregate button. We can rename any measure object we create. We can also perform grouping functions on these measures by selecting the conditions to each measure. For example Sum of Store sales or Maximum Store Cost etc... Grouping measure objects in one or more measures classes improves the organization of the universe. It also facilitates the end user’s ease of navigation when we click the Next button; Quick Design begins creating our universe.

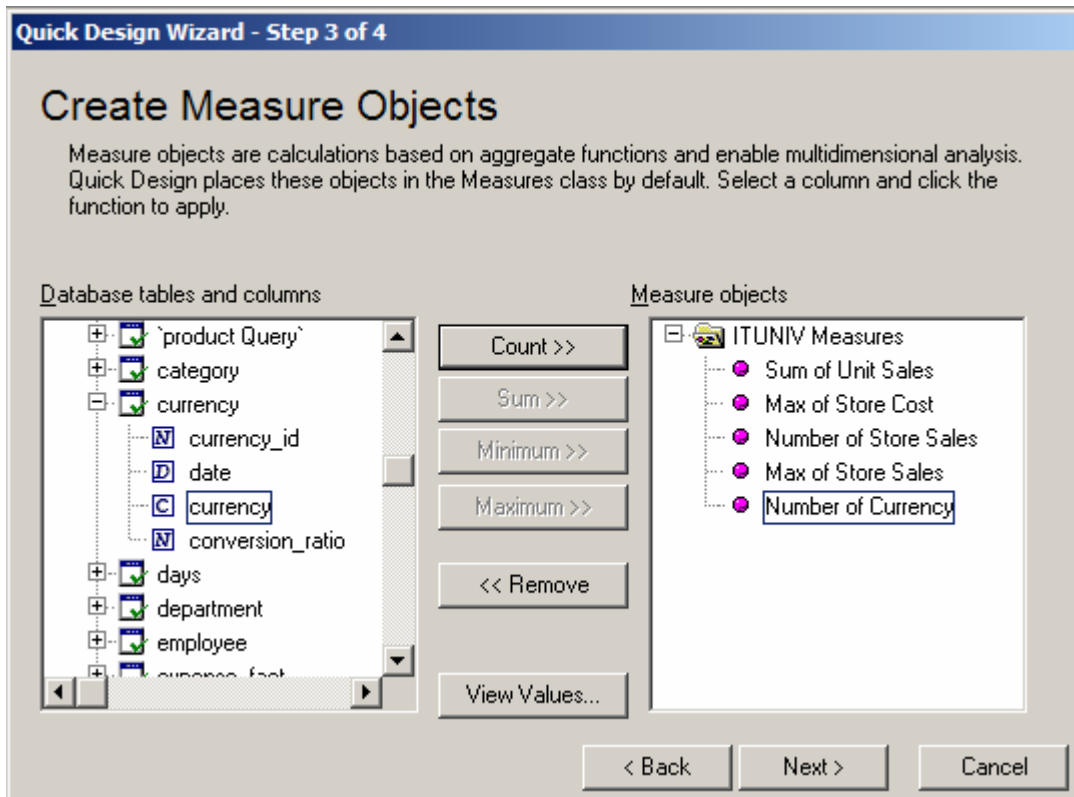


Figure 14: Creating Measure Objects

Generating the Universe: Quick Design automatically generates our new universe based on the parameters we specified. It indicates the number of classes, objects, and joins created in our universe. When we click the Finish button, DESIGNER displays the Universe pane and the Structure pane of our new universe.

Universe is a business-oriented mapping of the data structure found in databases, tables, columns, joins, etc. In the Business Objects User module, universes enable end users to build queries from which they can generate and perform analysis. Universes isolate end users from the complexities of the database structure as well as the intricacies of SQL syntax. In this sales universe we have 6 Dimension tables and one Fact table and 4 measures.

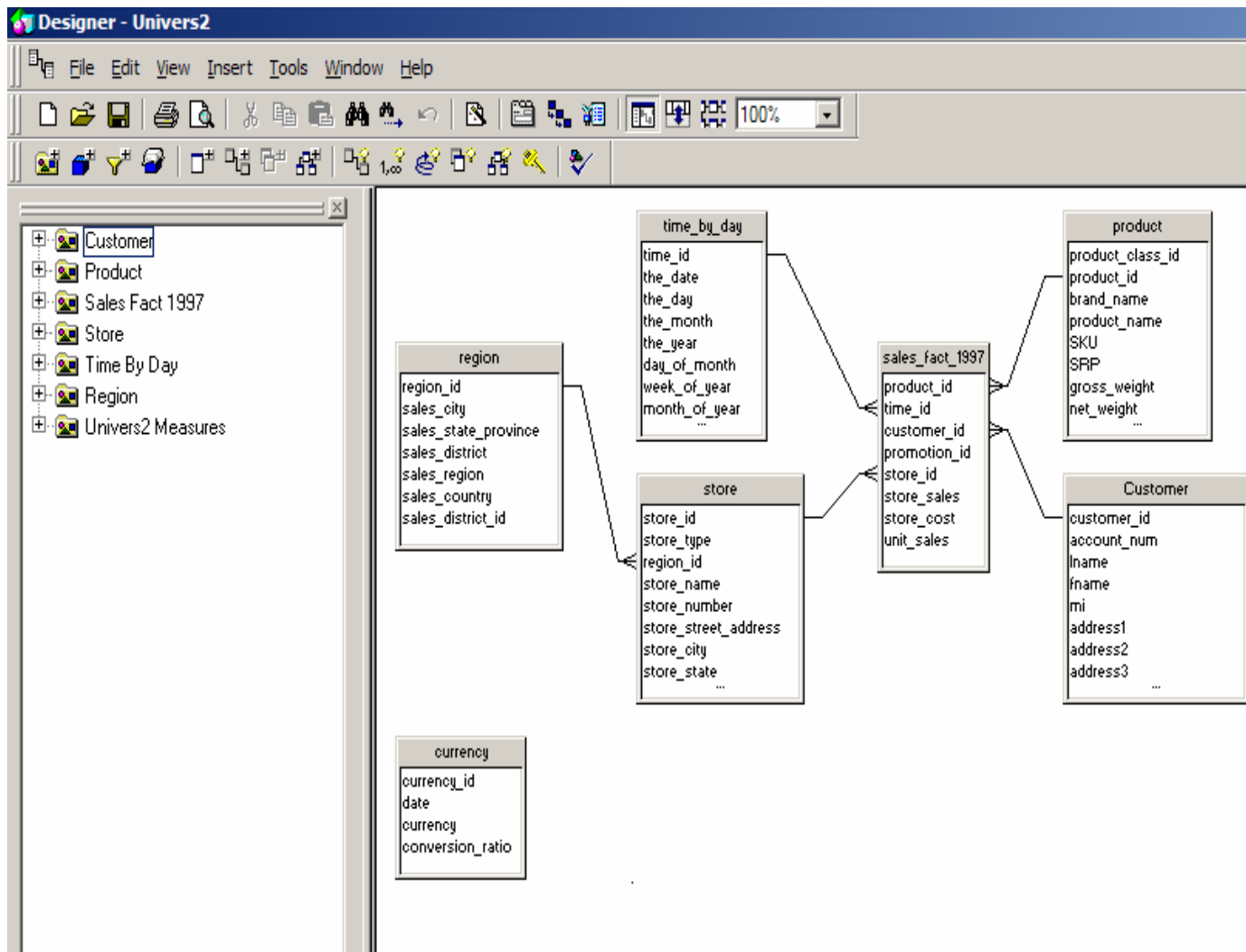


Figure 15: Sample Universe.

Saving a Universe: As with any other software application, we should remember to save our universes periodically throughout a work session. This dialog box also contains an option called Save for all users. To save all open universes, select the Save All command from the File menu, we should note that the universe name (the long name) and the file name by which we save it need not be the same. For the universe name we can enter up to 35 characters while for the file name we can enter up to 8 characters. When we save a universe, DESIGNER stores it as a file with an .unv extension. In Business Objects or Web Intelligence, the end user identifies the universe by the Universe name “.univ”.

Business Objects Report: Now we have Universe, by accessing the universe in to BO report we can analyze the data based on the requirement of the company. Once we have saved the universe we have started Business Objects for Reporting. A new wizard appears where we can create a new report. In this wizard we have 2 options; we can

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

create a report by using standard report format or we can use the template. To get the data for our report we can use the default universe we have created or we can specify how to access the data. After clicking the finish button we will go to the main page of Business Objects.

In the next step we get the Query Panel, the Query Panel is the one-step graphical interface that enables us to build queries. The query Panel is divided in to 3 parts 1.Classes and Objects 2. Result objects 3. Conditions. In the left hand side of Query Panel is Classes and Objects window. And Right hand side upper window is result objects and at the bottom of the window are result objects.

In the next step we select the required objects for reporting. We can define User Objects by applying some conditions and Rules (Formulas). To create a user object we must right click in the Classes and Objects window and select Create new object and give a specific name to that object for identification. By giving formula to that created object we can also define new measures by calculating the existed measures like “Profit = Sales- Store Cost”. We can apply some conditions to get our desired result. To get the result based on brand we have to apply some simple conditions on Brand Object. To apply condition, right click on brand Object and select apply simple condition and a new window appears in which we can see all the List of brand names. Select brands starting with Alphabet “A” we will get the details of all brands start with alphabet.

3.3 DATABASE

The Database taken for the purpose of Master Thesis is of Food chain supermarket. It is one of the pioneers in the food and beverage product sellers. The store has lot of varieties of brands and many products under each brand available in the store. Total number of different products sold by it is grouped into one category and again this has been divided into sub-categories basing on some condition. For example a product category called “Seafood”, it has been subdivided into Shellfish, Fresh Fish etc.

There are enormous number of records and information available in the database. The company’s sales figures for two years are available in the database. Each year has nearly 1, 00,000 records in it. This is what was requiring for the successful implementation of project with the help of different OLAP tools. This gave a scope to design Problem identification from the database and to solve the problem successfully. This in turn was helpful in making better and realistic decisions, faster and easier, which is the objective of thesis.

The database has so much of data in it. We have got the database of the company which uses Microsoft Access, by considering the output in mind and the design which can help in making any kind of data available for decision making.

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

3.4 PRACTICAL APPROACH

In our practical approach we start with

1. Reading installation guides of each tool.
2. Installing the reporting and analyzing software tools.
3. Start reading help documentation from the software CD and information from web site to know about the connection of database and how to generate reports.
4. After installing the tool connection with the database will be made.
5. Once the connection will be established, report generation process will begin followed by analysis process.

3.5 ANALYSES METHOD

After installing and generating reports in each tool, we can analyze which tool is more profitable for the company. The following factors will be correlated with results in order to conclude about the best available OLAP tool.

1. To check which tool can generate variety of graphs and reports and get detailed transactions from the database.
2. Analyzing target customers.
3. Multiple OS and language support.
4. Analyzing present market requirements
5. Percentage of usage.
6. Installation and troubleshooting problems
7. Scalability of the tool.
8. To check the ease of use and learn ability of each tool

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

4. Results

As the final step of the project, we have implemented our tools according to the method, in which these are some sample results in each tool.

4.1 *Cognos*

Cognos is world leader in business intelligence and performance planning software for the company. For implementing cognos we use Cognos Impromptu tool.

Cognos Impromptu delivers managed reporting for consistent and fast based decision making. It is a sophisticated report authoring tool to create wide variety of reports. It enables authors to create reports from any data source that can be deliver to report customers. Impromptu is used to create business context reports from simplest lists to most complex data grouping and can quickly design and delivered to the customers. Impromptu has framed based reporting interface which ease the report authors can authenticate virtually any report on the web.

During installation of Cognos we faced trouble and it was difficult to give appropriate parameters to installation manager during installation process without the help of installation guides. Database connection process was also a little bit complex. We generated different reports in cognos, it was difficult to learn cognos environment and to generate reports and graphs. Particularly wile generating graph we encountered a lot of technical problems. These problems were caused by lack of technical experience and difficult working environment.

A sample report generated with cognos is shown below.

Impromptu - [Report8]

File Edit View Insert Format Report Catalog Tools Window Help

Date: 2007-03-07

<Store Sales>

Product Id	Store Sales	Store Cost	Store Id
337	1,5000	0,5100	2
1512	1,6200	0,6318	2
963	2,4000	0,7200	2
181	2,7900	1,0323	2
1383	5,1800	2,1756	2
1306	7,4100	2,7417	2
1196	5,8400	1,9856	2
360	2,6200	1,0480	2
1242	3,9600	1,7028	2
154	1,9600	0,7252	2
483	3,8800	1,7072	2
77	5,6000	2,8000	2
533	4,8400	2,3232	2
310	0,7600	0,3420	2
1392	0,8300	0,3735	2
1303	1,3600	0,4352	2
748	4,4000	1,9360	2
1270	2,6900	1,0491	2
311	6,4500	2,3865	2
194	3,3600	1,5120	2
544	2,9500	1,0325	2
610	2,2700	0,7191	2

Catalog to use for reporting

100+

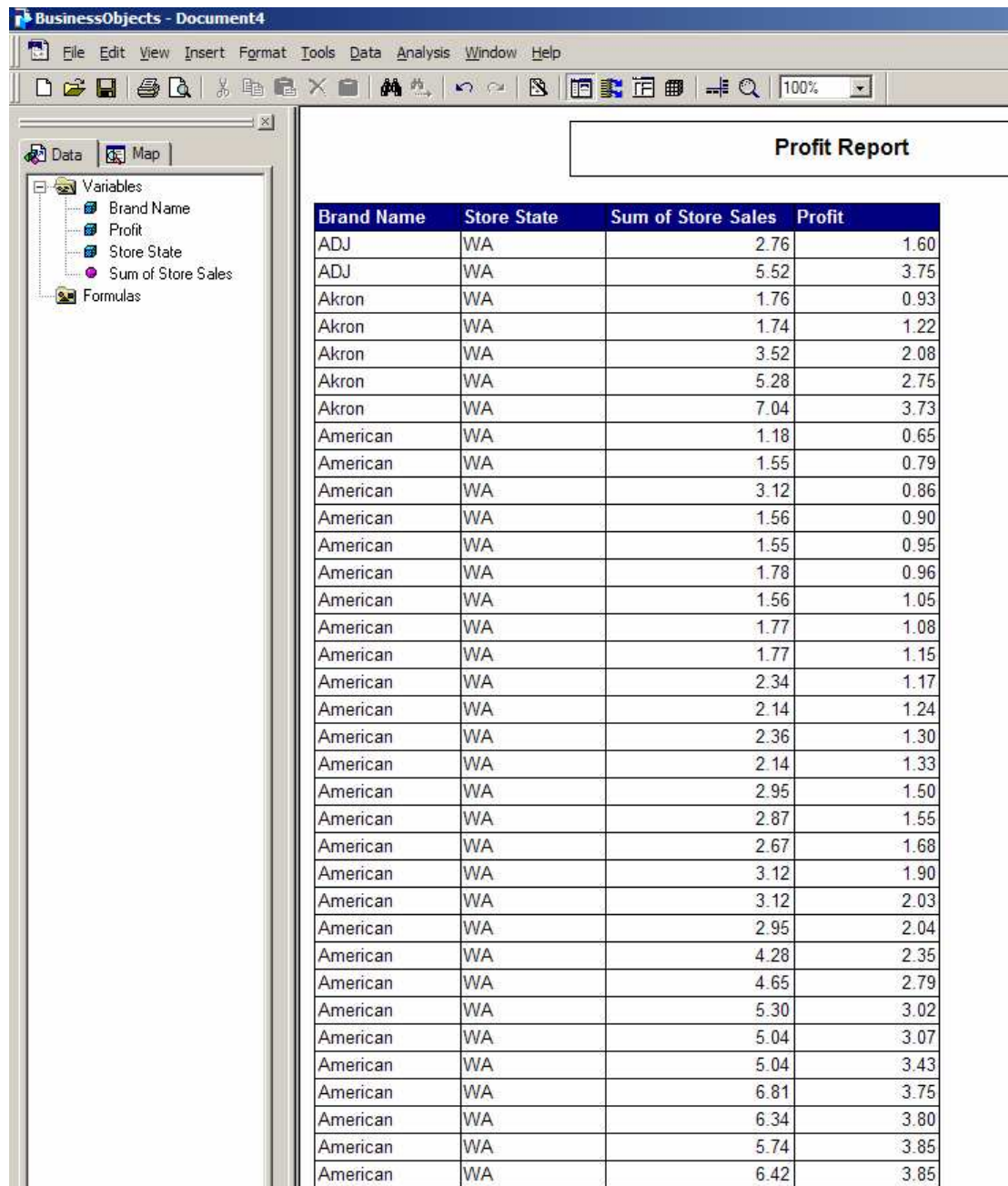
Figure 16: A Part of Sample Report Generated

4.2 Business Objects

It was fairly simple to install BO, and in the same way it went with learning the environment and generating reports graphs.

The configuration steps were easy going and self driven. Connection with database was also an easy process. More importantly the concepts behind report generation like Universe are close to natural phenomenon.

A sample report generated in BO is shown here.



BusinessObjects - Document4

File Edit View Insert Format Tools Data Analysis Window Help

100%

Profit Report

Brand Name	Store State	Sum of Store Sales	Profit
ADJ	WA	2.76	1.60
ADJ	WA	5.52	3.75
Akron	WA	1.76	0.93
Akron	WA	1.74	1.22
Akron	WA	3.52	2.08
Akron	WA	5.28	2.75
Akron	WA	7.04	3.73
American	WA	1.18	0.65
American	WA	1.55	0.79
American	WA	3.12	0.86
American	WA	1.56	0.90
American	WA	1.55	0.95
American	WA	1.78	0.96
American	WA	1.56	1.05
American	WA	1.77	1.08
American	WA	1.77	1.15
American	WA	2.34	1.17
American	WA	2.14	1.24
American	WA	2.36	1.30
American	WA	2.14	1.33
American	WA	2.95	1.50
American	WA	2.87	1.55
American	WA	2.67	1.68
American	WA	3.12	1.90
American	WA	3.12	2.03
American	WA	2.95	2.04
American	WA	4.28	2.35
American	WA	4.65	2.79
American	WA	5.30	3.02
American	WA	5.04	3.07
American	WA	5.04	3.43
American	WA	6.81	3.75
American	WA	6.34	3.80
American	WA	5.74	3.85
American	WA	6.42	3.85

Figure 17: Report in Business Objects.

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

5. Discussion

Business Intelligence is a powerful tool in a company's competitive arsenal, so perhaps it's not surprising that such software has been regarded as expensive to purchase and complicated to use. Business Intelligence software generally relies on a platform of medium-to-large scale Accounting or Enterprise Resource Planning (ERP) software to accurately capture detailed transactions. But recent innovations have broken through those barriers, making Business Intelligence practical for Small to Medium Enterprises and their Accountants.

To address the analysis needs of the sales department we examined the functionality of different OLAP technologies like Cognos and Business Objects. We found that Business Objects OLAP technology would be best suited for the recurring analysis performed by the sales department answering known questions from the summary to detailed data. While OLAP technologies seemed appropriated for the analysis needs we conclude that data mining technology would help the sales department find previously unknown patterns in detailed data, patterns that could prove to be of significant business value.

Unfortunately we were not able to experiment with Micro Strategy and SAS due to unavailability of those tools. These tools are also the most popular BI tools.

Cognos 8 Business intelligence has capabilities of Reporting, Analysis, Score carding, Dashboards, Business event management, Data integration. [13]

In reporting module, User interface supports more than 10 languages and reports are delivered in more than 25 languages and there is multiple export formats of reports. In this module, reports can be generated against any data source. Support is given for Windows, Linux, Unix and mixed platform deployments and Web based deployment.

Analysis module is a robust analysis tool, easy to use and easy to deploy. Score carding helps us to align teams and tactics with strategy; communicate goals consistently, and monitor performance against targets. It ensures ownership and accountability for performance. It makes it possible to Align decisions and tactics with own strategy and to execute against it.

Business dashboards communicate complex information quickly. Dashboards translate information from various corporate systems and data into visually rich presentations using gauges, maps, charts, and other graphical elements to show multiple results together. With dashboards we can have simple version of business issues at a glance, real world view with timely data refreshers and unite data from different core areas of the company.

Business event management tracks the significant events that need attention. The time to action and resolution is reduced by continuous monitoring of events and uses decision process and business process automation. For high performance business intelligence it is

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

necessary that data integration is done throughout enterprise wide to deliver data warehouses ready for business reporting and analysis.

Business Objects XI Release 2 allows non traditional users of BI tools to easily retrieve structured questions of business information. The new release includes Desktop intelligence which is the new name of Business Objects full client. By using this customers can protect their existing investment and benefit from the latest platform. The new features include support for Crystal Reports, Web Intelligence, OLAP Intelligence, Dashboard Manager, Desktop Intelligence, Performance Manager, Auditing, Publishing, Live office and Crystal Reports Explorer. It also include, integrated scheduling, full cross-platform support, auditing of all system interaction and support for Encyclopedia, Process Tracker, and Discussions. [14]

We used MS access database, to store data. Although there are other choices available for relational database management, but because database design and development was not in the scope of this thesis, we used a pre-developed sales database of a food chain in MS Access.

Following is a brief comparison of MS access database management system with Microsoft SQL Server and Oracle. Microsoft Access 2000 falls into the desktop category and works best for individuals and workgroups managing megabytes of data. In comparison with SQL Server 2000, Access uses file-server architecture, rather than client-server architecture. Access 2000 has many restrictions in comparison with SQL Server 2000 and cannot be used in the case of stable and efficient system with many concurrent users.

Some SQL Server 2000 and Access 2000 restrictions:

Feature	Access 2000	SQL Server 2000
SMP support	Not Supported	Supported
Tables	Relational tables	Relational tables, Temporary tables
Triggers	Not Supported	AFTER triggers, INSTEAD OF triggers
Procedures	Not Supported	Microsoft T-SQL statements
User-defined functions	Not Supported	Scalar functions, Inline table-valued functions, Multistatement table-valued functions
Views	Not Supported	Supported
Transaction logging	Not Supported	Supported
Recovery	Recovery to last backup	Recovery to last backup, recovery to the point of failure, recovery to a specific point in time
Integration with Windows NT security	Not Supported	Supported

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

Table 2: SQL Server 2000 and Access restrictions [15]

SQL Server 2000 and Access 2000 limits

Feature	SQL Server 2000	Access 2000
database size	1,048,516 TB	2 GB plus linked tables size
Objects in a database	2,147,483,647	32,768
Number of concurrent users	limited by available memory	255
Columns per table	1024	255
Table size	limited by available storage	1 GB
Number of tables in a query	256	32
Columns per SELECT statement	4096	255
nested subqueries	32	50
Number of enforced relationships	253	32
SQL statement size	65,536 * Network packet size (4 KB, by default)	approximately 64,000

Table 3: SQL Server and Access products limits [15]

SQL server Vs. Oracle

It is not true that SQL Server 2000 is better than Oracle 9i or vice versa. Both products can be used to build stable and efficient system. The stability and effectiveness of the applications and databases depend rather from the experience of the database developers and database administrator than from the database's provider. But SQL Server 2000 has some advantages in comparison with Oracle 9i and vice versa.

When compared to Oracle, SQL Server 2000 has some advantages:

SQL Server 2000 is cheaper to buy than Oracle 9i Database. It holds the top TPC-C performance and price/performance results. It is generally accepted as easier to install, use and manage.

The Oracle 9i Database advantages:

Oracle 9i Database supports all known platforms, not only the Windows-based platforms. PL/SQL is more powerful language than T-SQL.

More fine-tuning to the configuration can be done via start-up parameters. [15]

Since MS ACCESS is a relational database so the OLAP architecture we used is ROLAP. ROLAP and MOLAP architectures have major difference in the design philosophy. ROLAP architecture uses RDBMS for data storage and retrieval were as MOLAP architecture uses a multidimensional database. RDBMS supports various features that

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

enable quick and fast retrieval of data, and more over dynamic compilation of data. These include joins, star joins, complex predicates, parallel data queries etc. ROLAP systems is likely to be neutral in data compilation, it gives the system designer the capability to decide where to balance query and batch requirements, whereas MDDB are optimized to data retrieval and more focus on hashing and batch consolidation. The main goal of MDDB architect is to give maximum usage of query performance through pre-aggregation of data.

The schema used in database was Star schema. Star Schema is a relational database schema for multidimensional data. It is the simplest form of data warehouse schema that contains one or more dimensions and fact tables. Entity-relationship diagram between dimensions and fact tables resembles a star where one fact table is connected to multiple dimensions so it is called star schema. Star schema increases performance and ability to understand the data.

In [16] the same fact is described as “In OLAP, Snowflake schema approach increases the number of joins and poor performance in retrieval of data. In few organizations, they try to normalize the dimension tables to save space. Since dimension tables hold less space, Snowflake schema approach may be avoided.”

Star schema

Dimension table will not have any parent table

Hierarchies for the dimensions are stored in the dimensional table itself in star schema.

it is a highly de-normalized technique

Less number of joins

Snow Flake Schema

A dimension table will have one or more parent tables.

Hierarchies are broken into separate tables in snow flake schema.

If We apply normalized Principles to Star Schema Then It is Known As Snow Flake Schema.

Number of joins needed are greater than Star schema

Table 4: Comparison between star and Snow flake schema

The factors discussed in method section are applied in the following table. These factors are chosen from common factors used in analyzing usability, facilities (in terms of options offered to user) and from factors used to identify coverage/ wideness of software applications. All of the mentioned factors are selected from brochures used to market software systems, and a careful attention has been given to select factors to cover the whole aspects of a software system.

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

It has been observed by the authors that installation problems greatly impacts on user experience and retains the impression throughout the life of the software. On the other hand software which is easy to install gets user attention and confidence from the beginning. However from the above discussion, it should not be assumed that easy installation process of Business Objects have created a bias in our comparison.

Factor	Cognos	Business Objects
Installation problems	Rather difficult to install	Fairly easy
Choice of report and graph formats and styles	Offers rich choice of report formats and styles.	Rich collection of graph and report formats.
Target Customers	Small and medium scale	Small and Medium scale and independent vendors
Multiple OS support	Windows, Linux and Unix	Windows, HP-UX, Sun Solaris, AIX and Linux
Multiple language support	More than 10 languages for user interface, reports generated in more than 25 languages.	More than 14 languages.
System requirements	Ram 512 MB, Disk space 1 GB, web server, Database, Browser,	Ram 512 MB, Disk space 500 MB and above, Database.
Percentage of use	13	9
Learn ability	Not very good	Excellent and self driven
Scalability	Web support	Web support

Table 5: Factors applied to results.

As it is clear from the above table that, a lot of problems were faced by us during the initial phase (installation and configuration of tools). We had no experience with them before; all we did was to study the installation guides and troubleshooting web forums. Cognos was most difficult during and immediate after installation. A lot of information we had to provide to installation manager during installation process and after installation it was difficult to get started with the tool. We faced a lot of trouble to develop cubes which led to terminate our efforts to develop cubes in Cognos.

On the other hand during installation of Business Object we encountered no problems and it was an easy going process. More surprisingly the initial database setup was very easy

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

and the process of setting up environment was guided and self explanatory. It enabled us to setup and learn BO environment very quickly. Moreover it is very easy to change different settings in reports and changing graphs at run time

6. Conclusions

The authors had limited resources in terms of technical support, tools and time. Analyzing business intelligence tools performed in this thesis is an initial effort made to facilitate business and technical managers in deciding about to purchase OLAP tool.

Recommending the best available On-Line Analytical Processing (OLAP) tool was a difficult task for us initially. When only Cognos and BO are compared it is very easy to say that BO supersedes Cognos in most of the factors. However it cannot be regarded as a complete comparison, since two of the selected tools were not used.

This arises a strong future need of a thorough comparison of most commonly used OLAP tools.

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

7. References

- [1] IT Glossary, “Thomson Course Technology”, viewed at 11th Oct 2006, available at <http://www.course.com/careers/glossary/database.cfm>
- [2]. Planet Earth Projects, Inc, “The Purpose of Data Warehousing”, viewed at 11th Oct 2006, available at <http://www.p-e-p.com/DataWarehouse/Purpose.html>
- [3]. Abdourahamane FAYE, “From Data warehousing to Data Mining”, viewed at 25th Oct 2006, available at <http://www.codata.org/archives/2000/senegal-workshop/faye.htm>
- [4]. Paul Lane, Oracle 9i Data warehouse guide Release 2, “Extraction in Data Warehouses”, viewed at 25th Oct 2006, available at <http://www.lc.leidenuniv.nl/awcourse/oracle/server.920/a96520/extract.htm#11256>
- [5]. Leon Gong, Donna Venditti, DB2 Data Warehouse OLAP Services, “Part 1: Starting out with OLAP services released”, viewed at 31st Oct 2006, available at <http://www-128.ibm.com/developerworks/db2/library/techarticle/dm-0606gong/>
- [6]. Oracle 9i OLAP Users Guide, Release 2(9.2), “Why OLAP”, viewed at 5th Dec 2006, available at <http://saturn.uab.es/olap.920/a95295/basics2.htm#1005586>
- [7]. Oracle 9i OLAP Users Guide, Release 2(9.2), “Schema Modeling Techniques”, viewed at 5th Dec 2006, available at <http://www.lc.leidenuniv.nl/awcourse/oracle/server.920/a96520/schemas.htm>
- [8]. Cognos 2003, Cognos 7.0, Impromptu, Powerplay, Transformer, Help Documentation available in Installation CD
- [9]. Adrienne H. Slaughter, “OLAP”, <http://web.mit.edu/profit/PDFS/SlaughterA.pdf> , pp 1-6
- [10]. Business Objects 2002, Business Objects 5.1 Designer, Business Objects report, Help Documentation available in Installation CD, 2002
- [11]. A White Paper Prepared by Micro Strategy, “The Case for Relational OLAP” <http://www.insightstrategy.cz/solutions/archt/Case%20For%20ROLAP.pdf> , pp 5-9
- [12]. Corporate Performance Management, Overview, viewed at 17th Jan 2007, available at http://www.cognos.com/products/corporate_performance_management/index.html
- [13]. Reporting tool, “Cognos 8 Business Intelligence Reporting tool”, viewed at 2nd Mar 2007, available at <http://www.cognos.com/products/index.html>
- [14]. Reporting tool, “Business Objects XI Release 2 Reporting tool”, viewed at 1st Mar 2007, available at

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

http://www.businessobjects.com/pdf/products/businessobjects_xi_whats_new.pdf

[15]. Alexander Chigrik, "SQL Server 2000 vs Access 2000", viewed at 18th Feb 2007, available at http://www.mssqlcity.com/Articles/Compare/sql_server_vs_access.htm

[16]. Data Warehouse, "Star Schema", viewed at 1st Feb 2007, available at http://www.database-guru.com/dwh_starschema.php

[17] Leon Gong, Mike Olivas, Christine Poslosunzy, Donna Venditi, George McMillan, Deliver an effective and flexible data warehouse solution, "Part 3 : Design and Implement a warehouse ETL process, Aug 2005, available at <http://www-128.ibm.com/developerworks/db2/library/techarticle/dm-0508gong/>

[18]. Ralph Kimball and Margy Ross (2005), The Data Warehouse Toolkit: The Complete guide to Dimensional Modeling. 2nd Edition, John Wiley & Sons

[19]. Connolly T, Carolyn, Database Systems, second edition, Introduction to Database Systems, Addison Wesley Publishing Company; 2Rev Ed edition (August 1998)

[20]. E.F. Codd, The Relational Model for Database Management: Version 2, Addison-Wesley Pub. Co.; 1990

[21]. Humphries M, Hawkins M, Michelle C, Data warehouse architecture and Implementation, Prentice Hall, 1999

[22]. Erik Thomsen, OLAP Solutions: Building Multidimensional Information Systems, 2nd Edition, John Wiley & Sons, Inc. New York, 2002.

[23]. Phillip Blackwood, 11 Steps To Successful Data Warehousing. May, 2000. available at <http://advisor.com/Articles.nsf/aid/BLACP01>

Comparative Analysis of On-Line Analytical Processing Tools	Master Thesis
Sundeep Reddy Maddi and Vaseem Khan	

Comparative Analysis of On-Line Analytical Processing Tools

Sundeep Reddy Maddi and Vaseem Khan

© Sundeep Reddy Maddi and Vaseem Khan, 2007

Report no 2007:39

ISSN: 1651-4769

Department of Applied Information Technology

IT University of Göteborg

Göteborg University and Chalmers University of Technology

P O Box 8718

SE – 402 75 Göteborg

Sweden

Telephone + 46 (0)31-772 4895

[IT University of Göteborg]

Göteborg, Sweden 2007